Web Appendices to Accompany "Child Care Choices and Children's Cognitive Achievement: The Case of Single Mothers

Appendix A: Detailed Literature Review

This appendix gives a more detailed review of the literature than is provided in the main text.

A.1. The Effect of Maternal Employment and Child Care on Children's Cognitive Outcomes

Many prior studies, mostly in the developmental psychology literature, have used the NLSY to assess effects of maternal employment and childcare use on child cognitive development. Recent reviews of this literature include Love et al (1996), Blau (1999a), Lamb (1996), Haveman and Wolf (1994), Ruhm (2002) and Blau and Currie (2004). Less than half of these studies provide results that are interpretable in terms of effects of specific inputs.[1] Most present simple correlations between inputs and child outcomes and do not control for family and/or child characteristics. Furthermore, some of these studies use small samples, often nonrandomly selected. In most cases, no control for selection of children into childcare (or of mothers into employment) is implemented.[2]

Table 1 summarizes recent papers that use the NLSY data to assess effects of maternal employment on child cognitive outcomes. Clearly the evidence is inconclusive. Approximately a third of the studies report positive effects of maternal employment, a third report negative effects and the rest report effects that are either insignificant or that vary by the group studied or the timing of inputs. A similar picture is seen in Table 2, which summarizes recent papers that evaluate the effects of childcare (and/or childcare quality) on child outcomes.[3] Again, effects range from positive to negative, are often insignificant, and vary by group.

Reasons for this diversity of results may include the wide range of specifications that are estimated, and that many studies fail to control for endogeneity of employment and childcare. To make our exposition of the literature more clear, it is useful to have a specific framework in mind. Consider the following equation, interpretable as a cognitive ability production function:

$$\ln S_{ijt} = \alpha_1 T_{ijt} + \alpha_2 C_{ijt} + \alpha_3 G_{ijt} + \alpha_4 X_{ijt} + \mu_j + \delta_{ij} + \varepsilon_{ijt} \qquad (1)$$

Here S_{ijt} is a cognitive outcome (i.e., test score) for child i of mother j at age t. The log is typically taken as test scores are positive. T_{ijt} is a measure of the maternal time input up through age t. This might be a scalar, as in a cumulative specification, or a specification where only current inputs

[1] Some studies show associations between clusters of childcare arrangements/attributes and child development instead of assessing the impact of each input (Howes and Rubenstein (1985), Peterson and Peterson (1986), Studer (1992)). And in some cases, coefficient estimates or signs are not provided by authors (e.g., Howes and Rubenstein (1981)).
[2] See for example, Burchinal et al. (1995) and Parcel and Menaghan (1990).
[3] Since the literature contains fewer studies of childcare, Table 2 is not restricted to studies that use NLSY data only.

matter, or a vector, if inputs at different ages are allowed to have different effects. Similarly, C_{ijt} is a measure of nonmaternal time inputs (i.e., childcare), and G_{ijt} represents goods and services used in the production of child ability. Next, X_{ijt} is a set of controls for the child's initial skill endowment. This may include variables such as mother's age, education, AFQT score, etc. (meant to capture the inherited ability endowment), and initial characteristics of the child such as gender, race and birthweight. Turning to the error components, μ_j and δ_{ij} are family and child effects, which capture parts of the *unobserved* skill endowment of the child. Finally, ε_{ijt} is a transitory error term that may be interpreted as measurement error inherent in the test plus error in recording test results, along with shocks to the cognitive development path.

While this general setup seems to underlie, at least implicitly, most papers in the literature, none actually estimate equation (1), and many estimate equations that seem quite far from it. One fundamental problem is that the maternal time input *T* and the goods inputs *G* are not directly observed. Most papers ignore the problem that *T* is unobserved, simply using maternal employment or childcare use in place of maternal time.[4]

Similarly, most papers simply ignore *G,* while a few proxy for it using household income or the NLSY's "HOME" environment index. The latter is problematic as it is based not just on goods inputs (e.g., books in the home) but also on maternal time inputs (e.g., time reading to the child). Baydar and Brooks-Gunn (1991) estimate effects of both maternal employment and childcare, but do not include goods/services. Desai et al. (1989) use maternal employment to proxy for *T*, average number of child care arrangements during the first three years after childbirth to proxy for *C* and household income to proxy for *G*. But, as noted by Rosenzweig and Schultz (1983), Rosenzweig and Wolpin (1994) and Todd and Wolpin (2003, 2007), it can be difficult to interpret production function estimates when proxies are used for key inputs. To our knowledge, only James-Burdumy (2005) discusses the relationship between her estimating equation and a child ability production function by pointing out the difficulty of interpreting estimates when proxies are used for maternal time and goods inputs. We discuss this issue in more detail in Section 4 and in Web Appendix B.

Secondly, most papers in the literature have estimated specifications that include only *current* inputs. This is a strong assumption, especially in light of the tradition in the human capital literature of letting cumulative inputs matter. One could think of the effect of inputs cumulating over time or having a more general specification according to which the whole history of inputs since

[4] Also, most papers use one or the other of these variables, and do not examine both. For example, Vandell and Ramanan (1992) estimate the effect of maternal employment on child's cognitive outcomes but do not include childcare time as an additional input, while Caughy et al (1994) do the reverse.

childbirth matters for the child's outcome at time *t*. Most papers do not discuss the implications of their assumptions regarding timing of inputs.[5] We also discuss this issue in Section 4, and test for the importance of lagged inputs in Section 6.2 and Web Appendix D.

Finally, most papers estimate equation (1) by OLS, ignoring potential endogeneity of inputs – i.e., potential correlation of maternal work and childcare use decisions, and goods inputs, with the unobserved ability endowments μ_j and δ_{ij}. A few recent studies try to overcome this problem by either: (1) using a very extensive set of variables to proxy for unmeasured endowments, (2) using child or family fixed effects, or "value added" models,[6] and/or (3) using instrumental variables.

Consider first the studies that can be classified as using extensive controls for the child's skill endowment. Among others, Han et al (2001), Baydar and Brooks-Gunn (1991), Parcel and Menaghan (1994), Vandell and Ramanan (1992) and Ruhm (2002), use an extensive set of observable characteristics of the child and the mother, including mother's AFQT score. In spite of this, the results of these papers are inconclusive. For example, Ruhm (2002) finds significant *negative* effects of maternal employment on math scores while Parcel and Menaghan (1994) report small *positive* effects of maternal employment on child cognitive outcomes. Baydar and Brooks-Gunn (1991) find that maternal employment in the child's first year *negatively* affects cognitive outcomes, while Vandell and Ramanan (1992) find *positive* effects of early maternal employment on math achievement, and of current maternal employment on reading achievement.

Next, consider the studies that use fixed effects. Chase-Lansdale et al. (2003) use child fixed effects models to assess the effect of maternal employment on child outcomes. They analyzed 2,402 low-income families during the recent era of welfare reform. Their results suggest that mothers' transitions off welfare and into employment did not cause negative outcomes for preschoolers. They note, however, that this approach does not account for endogeneity of these transitions, and they do not attempt to use welfare rules as instruments for maternal employment as we do here.

James-Burdumy (2005) estimated household FE models using 498 sibling children in the NLSY. Her results suggest that effects of maternal employment vary depending on the particular cognitive ability assessment used and the timing of employment.[7] The use of sibling differences

[5] Notable exceptions are Blau (1999a) and Duncan (2003). Some papers use maternal employment (or childcare use) at different years after childbirth but do not discuss implications of their choice in terms of properties of the underlying production function (e.g., Waldfogel et al. (2002), Vandell and Ramanan (1992), and Baydar and Brooks-Gunn (1991)).

[6] In the value-added approach, the test score in period *t* (S_{ijt}) is a function of the outcome in period *t-1* and the inputs in period *t*, the idea being that the lagged test score proxies for the child's ability at the start of a period.

[7] According to James-Burdumy (2005)'s fixed effects (FE) estimates in her Table 5, an increase in maternal work from 0 to 2000 hours in year 1 of a child's life reduces the PIAT math score (measured at ages 3 to 5) by (-.00117)×2000=-2.34

eliminates the mother (or household) fixed effect μ_j from (1) but does not eliminate the child fixed effect δ_{ij}. It is plausible that mothers make time compensations for children depending on their ability type. Using household fixed effects does not solve this problem, as maternal employment is then correlated with the sibling specific part of the cognitive ability endowment. In addition, the FE estimator requires that input choices are unresponsive to prior sibling outcomes. If inputs to child *i'* are responsive to outcomes for child *i*, then ε_{ijt} will be correlated with those inputs.

Blau (1999a) and Duncan and NICHD (2003) both study the effects of childcare use and quality on child outcomes. They use similar methodologies, including a wide range of proxies for unmeasured child ability (e.g., mother's AFQT and education), controls for many aspects of the home environment, and use of fixed effects and value added specifications. The main difference is that Blau (1999a) uses the NLSY while Duncan uses the NICHD Study of Early Child Care. Blau (1999a) concludes "child care inputs … during the first three years of life have little impact on … child outcomes …" while Duncan finds modest positive effects of improved child care quality.

From our perspective, a key difficulty in interpreting the Blau and Duncan results is that their specifications don't allow one to estimate the effect of maternal time *per se*. Both studies include the HOME environment index, which contains both goods inputs, like books in the home, and also time inputs, like how often the child is read to, eats meals with parents, or talks with the mother while she does housework. Thus, the coefficients on maternal work or childcare capture the effects of those variables holding HOME <u>fixed</u>, which means holding some maternal time inputs fixed. In contrast, we are interested in the total impact of maternal time on child outcomes, including how a decline in the time input (from increased work or childcare use) affects time reading to the child and so on.

Finally, Currie and Thomas (1995) use the NLSY to look specifically at how pre-school affects outcomes. Using sibling differences and extensive controls for ability endowments, they estimate a year of Head Start increases PPVT scores by roughly 7%, while other types of pre-school have no effect. The Head Start effect persists for whites, but is wiped out by age 10 for blacks.

The Blau, Duncan-NICHD and Currie-Thomas papers all contain useful discussions of the limitations of fixed effects and value added specifications. As they point out, neither approach is a panacea for dealing with unobserved child ability, as each relies on assumptions that can be stronger than OLS. For example, the household FE estimator requires that input choices be unresponsive to the child specific part of the ability endowment. The value added model faces the problem that

---

points. This is similar to the effect we estimate for one year of full-time work (-2.1%). But she finds no significant effect of maternal employment after the first year. In contrast, we find maternal time is just as important in years 2+.

estimates of lagged dependent variable models are inconsistent in the presence of fixed effects like μ_j and δ_{ij}.[8] Neither approach, nor child fixed effects, deals with endogeneity arising because current inputs may respond to lagged test scores. An IV approach is needed to deal with this problem.

To our knowledge, only two prior papers have attempted to use IV in this context. These are Blau and Grossberg (1992) and James-Burdumy (2005).[9] Both look at effects of maternal work on child outcomes, and do not examine effects of childcare use *per se*. More importantly, both papers suffer from the problem that the instruments are extremely weak. As a result, the standard errors on the maternal work variables in their 2SLS regression are so large that no plausibly sized effect could possibly be significant (i.e., in each case, to attain 5% significance, maternal work over a three year period would have to change a child's tests scores by roughly 50 points or 3 standard deviations).[10] Thus, we would argue that their attempts to implement IV were not successful. Similarly, Currie and Thomas (1995) report they attempted to use IV but could not find sufficiently powerful instruments.

The main advantage of our approach is that the welfare policy and local demand instruments that we employ are much stronger. Indeed, the first stage marginal R^2 values we obtain using these instruments (i.e., about .09) are fairly large, and, in the second stage, the standard error on childcare does not "explode" when these instruments are used.

Bernal (2006) takes a different approach by estimating a structural model of work and child care choices of *married* women. She estimates a child cognitive ability production function – which includes mother's work and childcare use as inputs – jointly with the mother's work and childcare decision rules, thus implementing a selection correction. Her results suggest rather sizable effects of maternal employment and childcare use on child cognitive ability. In particular, one full year of maternal work and childcare use causes a 1.8% reduction in test scores of children ages 3 - 7.[11]

It is interesting to extend this work to *single* mothers for several reasons. The first is to see if results generalize. Second, single mothers are of special policy relevance, as welfare reform led to

[8] Estimation of a first-differenced version of the value-added specification would eliminate the fixed effects, but Blau (1999a) points out that this is difficult or impossible due to limitations of existing data. This would require three outcome observations and two lagged input observations. Even if feasible, this approach would entail a severe efficiency loss.

[9] James-Burdumy's preferred specification uses sibling differences to control for household FE, and does not use IV. However, she notes that maternal employment may be endogenous in the differenced equation, due to correlation of the time-varying parts of the errors in the child outcome and maternal employment equations. Of course, another source of endogeneity is correlation between unobserved child ability and the mother's decisions about work and childcare.

[10] For Blau and Grossberg (1992), who use work experience prior to childbirth to instrument for maternal employment, compare columns 1 and 2 of their Table 2. For James-Burdumy (2005), who uses the percentage of the county labor force employed in services to instrument for maternal employment, compare columns FE and IV-FE from her Table 3.

[11] Liu et al. (2003) also adopt a structural approach to estimate effects of maternal employment and school inputs on test score outcomes for 5 to 15 year olds in the NLSY. They also find a negative effect of maternal employment on child outcomes. Obviously, the focus in Bernal (2006) and here is rather different, as we are interested in pre-school inputs.

large increases in their work/childcare use. Third, welfare rules have large effects on work/childcare use by single mothers, so as instruments they provide a strong basis for identification. It is difficult to find plausibly exogenous variables that impact behavior of married women so strongly.

A.2. The Relationship between Test Scores and Subsequent Outcomes (Wages, Education, etc.)

Several studies have examined the relationship between test scores as early as age 7 and subsequent outcomes like educational attainment and wages. This research finds that measures of cognitive achievement recorded in childhood are strong *predictors* of a variety of outcomes later in life. This highlights the importance of understanding what determines ability of individuals at early stages of life, particularly for the design of public policy aimed at improving labor market outcomes. We summarize some of these studies in this section.

First, consider studies using U.S. data. In the NLSY, Neal and Johnson (1996) find that scores at ages 14 to 21 on the Armed Forces Qualifying Test (AFQT), an IQ-type measure, are highly significant predictors of wages at ages 26 to 29. Murnane, Willett and Levy (1995) use two longitudinal surveys of high school seniors to document a strong relationship between their math test scores and wages at age 24. Zax and Rees (1998) use the Wisconsin Longitudinal Study (WLS) to document that age 17 IQ is a strong predictor of wages at ages 35 and 53.

The studies linking test scores at the earliest ages to later outcomes use the British National Child Development Study (NCDS). Hutchinson, Prosser and Wedge (1979) use the NCDS to link test scores at age 7 with scores at age 16. Similarly, Connolly, Micklewright and Nickell (1992) find a significant positive relationship between test scores at age 7 and earnings at age 23 (in a sample of young men who left school at age 16). More recently, Robertson and Symons (1996) and Harmon and Walker (1998) find a positive association between age 7 test scores and earnings at age 33. And Currie and Thomas (2001) show that a one standard deviation increase in age 16 math scores is associated with a 14% higher wage rate and a 7% higher employment rate at age 33 (for low or medium-SES individuals). In addition, they provide evidence that age 7 (math) test scores are strong predictors of age-16 math test scores.

From our perspective, a limitation of these studies is they all look at test scores at age 7 or older (14 or older in the U.S. case). Do tests scores at even earlier ages predict later achievement? In Appendix 1 we present evidence from the NLSY that PPVT scores at age 4 and PIAT reading and math scores at ages 5 to 6 are significantly correlated with educational attainment of youth who are at least 18 years old. For example, consider a one-point increase in the math score at age 6 (i.e., roughly a 1% increase, as the mean score is 99.7). Holding parental background variables like mother's education fixed, this is associated with increased educational attainment (measured at age 18 or later) of approximately .019 years. Similarly, a one-point (roughly 1%) increase in the reading score at age 6 is associated with an increase in highest grade completed of approximately .025 years. These estimated impacts are fairly substantial. For example, our estimates imply a year of full-time maternal work and informal childcare use reduces test scores by roughly 2.6%. This translates into an effect on completed schooling of roughly .050 to .065 years, a large effect.[12]

A striking aspect of the Appendix 1 results is that mother's AFQT score is not a significant predictor of completed education. Thus child test scores, even at ages 4-6, are better predictors of later outcomes than mother's scores. For example, in the equation that that includes the child's age 6 PIAT-math score, a one standard deviation increase in the math score is predicted to raise completed education by (.0191)·(11.7)=.223 years. In contrast, a one standard deviation increase in mother's AFQT score is predicted to increase completed education by only (.00128)·(18.3)=.023 years. So the point estimates imply an effect of the child's math score ten times greater than that of the mother's AFQT score. And in the equation that includes the child's age 6 PIAT-reading score, the coefficient on mother's AFQT is essentially zero.[13]

Appendix B: Interpreting Estimates of the Child's Cognitive Ability Production Function

As with any estimation method, the interpretation of our IV estimates of the effect of child care on child cognitive outcomes depends on a number of maintained assumptions. In this appendix we provide a more detailed discussion of those assumptions, and discuss how their violation could lead to difficulties in interpreting the estimates.

In the human capital production framework (see Ben-Porath (1967)) current and past inputs interact with an individual's genetic ability endowment to generate human capital. Leibowitz (1974) first used this framework to examine how investments in children add to preschool stocks of human capital. The acquisition of preschool human capital is analogous to the acquisition of human capital through schooling or on-the-job training, except that, at preschool ages, inputs are generated by *joint* parental/child decisions (e.g., child tastes presumably affect parent input choices), not by choices of the child alone. Here, we focus on the cognitive ability component of human capital.

[12] The following back-of-the-envelop calculation helps put these figures in perspective: Say people are of two types, those who finish high school (12 years) and those who finish college (16 years), and that 20% finish college. To increase average completed schooling by .06 years, the percentage finishing college must increase to 21.5%, a 7.5% increase.

[13] At earlier ages the point estimates still imply larger effects of the child score than the mother's score, although the difference is not as great. For example, at age 5, a one standard deviation increase in the child's PIAT-R score is predicted to increase completed education by (.0096)·(15.3)=.147 years, while a one standard deviation increase in the mother's AFQT is predicted to increase completed education by (.0045)·(18.3)=.082 years. Similarly, in the equation that contains the child's PIAT-M score at age 5, the analogous figures are .091 vs. .069 years.

Let A_{it} be child i's cognitive ability t periods after birth. We write a production function:

(2) $$\ln A_{it} = A(\widetilde{T}_{it}, \widetilde{G}_{it}, \widetilde{C}_{it}, \omega_i)$$

where $\tilde{T}_{it}$, $\tilde{G}_{it}$ and $\tilde{C}_{it}$ are vectors of period-by-period inputs of maternal time, goods and childcare time, respectively, up through period t, and ω_i is the child's ability endowment. Goods inputs may include nutrition, books and toys that enhance cognitive development, etc.. Childcare inputs capture contributions of alternative care providers' time to child cognitive development. These may be more or less effective than mother's own time. E.g., care in a group setting may contribute to child development by stimulating interaction with other children, learning activities at pre-school, etc..

Several difficult issues arise in estimation of (2). First, estimation of a completely general specification, where inputs may have a different effect at each age t, and where the endowment ω_i may differentially affect ability at each age, is infeasible due to proliferation of parameters.[14] Thus, we obviously need to restrict how inputs enter (2).

One simplification, familiar from the human capital literature, is to assume that: (i) only cumulative inputs matter, rather than their timing, and (ii) the effect of the permanent unobservable is constant over time (e.g., in a Mincer earnings function, only cumulative education and experience affect human capital, and the unobserved skill endowment has a constant effect). We first consider a specialization of (2) that adopts these assumptions, and consider some feasible relaxations later. Letting $\widehat{X}_{it} = \sum_{\tau=1,t} X_{it}$ be the cumulative amount of input X up through time t, and assuming that cumulative inputs affect $ln\ A_{it}$ linearly, we obtain a special case of (2) that takes the form:

(3) $$\ln A_{it} = \alpha_0 + \alpha_1 \widehat{T}_{it} + \alpha_2 \widehat{C}_{it} + \alpha_3 \ln \widehat{G}_{it} + \omega_i$$

We now consider problems of estimating the production function in the special case of (3).[15]

The second issue we face is the selection (or endogeneity) problem that arises because inputs may be correlated with the child ability endowment ω_i. To clarify this problem, assume the ability endowment is given by the equation:

(4) $$\omega_i = \beta_0 + \beta_1 Z_i + \hat{\omega}_i ,$$

where Z_i is a vector of mother/child characteristics correlated with the child ability endowment (e.g., mother's education and AFQT score, child gender and birthweight), and $\hat{\omega}_i$ is the part of the ability

---

[14] For instance, if the effect of just one input is allowed to differ between every pair of input and output periods t and t', and we examine outcomes for 20 quarters after birth, we obtain 20·21/2=210 parameters for that input alone.

[15] Letting cumulative goods enter in log form is analytically convenient, for reasons that will become apparent later.

endowment that is mean independent of observed mother and child characteristics. Next, as an illustration, assume a mother's decision rule for childcare time at time t, C_{it}, can be written:

$$C_{it} = \pi_0 + \pi_1 Z_i + \pi_2 \hat{\omega}_i + \pi_3 cc + \pi_4 R_{it} + \varepsilon_{it}^c ,$$

where cc is the price of childcare (assumed constant),[16] R_{it} is a set of welfare rules facing the mother at time t, and ε_{it}^c is a stochastic term subsuming tastes for childcare use (both permanent and transitory taste shocks), and shocks to childcare availability and the mother's offered wage rate. The presence of $\hat{\omega}_i$ in the decision rule means $\widehat{C}_{it}$ is endogenous in (3), and we will require instruments that affect C_{it} yet are uncorrelated with $\hat{\omega}_i$ and ε_{it}^c. Below we argue the welfare rules R_{it} can plausibly play this role.

The third key issue in estimating (3) is measurement of maternal time and goods inputs. One can imagine a model where mothers decide how much "quality" time to devote to the child while at home (e.g., children's time is divided between childcare, "quality" time with the mother, and time spent watching TV while she does housework). But, we don't observe actual contact time between mothers and children, let alone how much is "quality" time. So, as is typical in the literature as a whole, we simply side-step the issue by assuming that $T_{it} = T\text{-}C_{it}$, where T is total time in a period. Thus, we distinguish between only two types of time (i.e., time with the mother and time in childcare). Then, we can rewrite (3) as:

$$\text{(5)} \qquad \ln A_{it} = \alpha_0 + (\alpha_1 T)\cdot t + (\alpha_2 - \alpha_1)\widehat{C}_{it} + \alpha_3 \ln \widehat{G}_{it} + \omega_i$$

Thus, we can only estimate $\alpha_2 - \alpha_1$, the effect of time in childcare *relative* to that of mother's time.

An issue we abstract from here is that maternal work time may influence how much of $T\text{-}C_{it}$ is "quality time." For example, a mother who uses childcare but does not work might devote more of $T\text{-}C_{it}$ to "quality time." Thus, *maternal work time might enter the production function directly*, independently of how it affects the goods input (via the budget constraint) or how it affects C_{it}. However, for single mothers it is very difficult to address this issue, because child care and maternal work time are extremely highly correlated (ρ=.94).[17] Thus, attempts to include both in the model fail due to severe colinearity.

---

[16] That the price of childcare cc is assumed constant over mothers/time is not an accident. A key problem confronting the literature on childcare is that the geographic variation in cc seems too modest to use it as an IV for childcare usage.

[17] Obviously, single mothers must use childcare to work, and most cannot afford day care otherwise. In contrast, for married women, use of childcare while not working is fairly common (see Bernal (2006)).

The fourth key issue in estimation of (3) is that goods inputs G_{it} are largely unobserved. For example, the NLSY contains information on books in the home, but not nutrition, health care, tutors, recreation, etc.. This problem of missing inputs plagues the entire literature, not just our study (see, e.g., Todd and Wolpin (2007) for a discussion).

To deal with this missing input problem, consider a decision rule for the cumulative goods input into the child's ability, conditional on observed mother/child characteristics Z_i (which affect permanent income and preferences), the child's ability endowment $\hat{\omega}_i$,[18] cumulative income since childbirth, child age (t), and childcare usage decisions, given by:

$$\text{(6)} \qquad \ln \widehat{G}_{it} = \gamma_0 + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \widehat{I}_{it}(W, H; R) + \gamma_4 t + \gamma_5 \widehat{C}_{it} + \varepsilon_i^g .$$

where the stochastic term ε_i^g captures the mother's taste for investment in the form of goods.[19] The notation $\widehat{I}_{it}(W, H; R)$ highlights the dependence of income on wages, hours of market work, and the welfare rules R that determine how benefits depend on income.

Equation (6) can be interpreted as a conditional decision rule or demand function, obtained in stage two of an optimization process, where, in stage one, a mother chooses childcare time C and hours of market work H, and in stage two chooses G. [Note that the temporal aspect here is purely artificial, as in a two-stage budgeting solution]. Equation (6) can also be interpreted as a linear approximation to the decision rule that would be generated by several alternative models of investment, both static and dynamic.[20] The key thing captured by (6) is that a mother's decisions about goods inputs into child development may be influenced by (i.e., made jointly with) her decisions about hours of work and childcare. Substituting (6) and (4) into (5) we obtain:

$$\text{(7)} \qquad \begin{aligned} \ln A_{it} &= \alpha_0 + (\alpha_1 T) \cdot t + (\alpha_2 - \alpha_1) \widehat{C}_{it} \\ &\quad + \alpha_3 [\gamma_0 + \gamma_1 Z_i + \gamma_2 \hat{\omega}_i + \gamma_3 \ln \widehat{I}_{it} + \gamma_4 t + \gamma_5 \widehat{C}_{it} + \varepsilon_i^g] + \omega_i \\ &= (\alpha_0 + \alpha_3 \gamma_0) + (\alpha_1 T + \alpha_3 \gamma_4) \cdot t + (\alpha_2 - \alpha_1 + \alpha_3 \gamma_5) \widehat{C}_{it} \\ &\quad + \alpha_3 \gamma_3 \ln \widehat{I}_{it} + \alpha_3 \gamma_1 Z_i + (1 + \alpha_3 \gamma_2) \hat{\omega}_i + \alpha_3 \varepsilon_i^g \\ &= \varphi_0 + \varphi_1 \cdot t + \varphi_2 \widehat{C}_{it} + \varphi_3 \ln \widehat{I}_{it} + \varphi_4 E_i + \hat{\hat{\omega}}_i + \hat{\varepsilon}_i^g \end{aligned}$$

[18] Note that the child's ability endowment may matter for two reasons: Either mother's may choose good inputs based on the child's ability (e.g., they may buy educational toys to compensate a child who is having certain learning problems) or because child ability affects the types of inputs a child demands (e.g., a high ability child may request more books).

[19] This would arise due to heterogeneous preferences for child quality. ε_{it}^g may also be influenced by the child's tastes.

[20] For example, if γ_3=0 there is a fixed rate of investment determined by <u>permanent</u> characteristics, and the cumulative goods input grows at a rate given by γ_4. At the other extreme, if γ_1=γ_2=γ_4=0 and γ_3=1 then demand for goods is simply proportional to <u>current</u> income (G_{it}=exp(γ_0)·I_{it}).

Equation (7) is estimable, because all the independent variables are observable. However, we must be careful about the appropriate estimation method and interpretation of the estimates. As we have noted, childcare utilization may be correlated with the unobserved part of the child ability endowment $\hat{\omega}_i$. Furthermore, childcare use may be correlated with $\hat{\varepsilon}_i^g$, the unobserved taste shifter in equation (6), if tastes for childcare use ε_{it}^c are correlated with tastes for goods investment.[21] Then, estimation of (7) using OLS is not appropriate.

To our knowledge, it has not been previously noted that consistent estimation of an equation like (7) requires instruments that are not only uncorrelated with the unobserved part of the child's skill endowment, $\hat{\omega}_i$, but also with the mother's tastes for goods investment in the child, ε_i^g. In order for the welfare rule parameters R_{it} to be valid instruments for cumulative childcare in estimating (7), they must be uncorrelated with these two error components. This seems like a plausible exogeneity assumption.[22] We would make a similar argument for local demand conditions.

The cumulative income variable in (7) is also potentially endogenous, for multiple reasons. First, income depends on the jointly made childcare use and work decisions. Hence it is potentially correlated with child ability for the same reasons as were operative for childcare usage. Second, income depends on the mother's wage rate, which depends on her ability endowment. To the extent that this ability endowment is not perfectly captured by Z_i, and the residual part is correlated with the child ability endowment, this will also generate correlation between the mother's income and $\hat{\omega}_i$. Thus, we need to instrument for mother's income as well. Again, we will argue that welfare rules R_{it} and local demand conditions provide plausibly valid instruments, as they have important effects on work decisions, yet are plausibly uncorrelated with child ability endowments.

Assuming that instrumental variables provides consistent estimates of (7), it is important to recognize that the childcare "effect" that is estimated is $\beta_2 = \alpha_2 - \alpha_1 + \alpha_3 \cdot \gamma_5$. This is the effect of child care time (α_2) *relative* to the effect of mother's time (α_1) plus the effect of any change in goods inputs that the mother may choose as a result of using childcare ($\alpha_3 \cdot \gamma_5$). In the main text we assume that $\gamma_5 = 0$, so that the estimated childcare effect is in fact $\alpha_2 - \alpha_1$, the effect of childcare time relative

[21] For instance, a mother with a high taste for child quality may both spend more time with the child (i.e., use less day care) and invest more in the child in the form of goods. This would tend to bias estimated effects of childcare usage in a negative direction, since not only the maternal time input but also the goods input is lower for children in childcare.
[22] In Web Appendix G we report means of child test scores prior to 1990 by State, broken down by whether the State subsequently implemented welfare waivers (i.e., moved towards Welfare reform early), and whether the State implemented strict or lenient welfare rules after 1996. There is no significant difference in average pre-reform test scores between "strict" and "lenient" States.

to maternal time. This is equivalent to assuming that childcare time does not alter the marginal utility of investment in child quality via goods, so that C drops out of (6). Here we want to consider the implications of failure of this assumption.

If $\gamma_5 \neq 0$, it leads to some important limitations for IV estimates of (7). First, there is the issue that interpretation is more subtle. We must always bear in mind that we are estimating $\alpha_2 – \alpha_1 + \alpha_3 \cdot \gamma_5$, which includes the behavioral response term $\alpha_3 \cdot \gamma_5$, and thus we are not estimating the pure technological effect of childcare.

Second, there is the issue that γ_5 is a reduced form parameter. Thus, even given a consistent estimate of $\beta_2 = \alpha_2 – \alpha_1 + \alpha_3 \cdot \gamma_5$, we must be careful to use it only to evaluate the impact of changes in childcare induced by policy changes that leave γ_5 unchanged. Some policies will do this while others will not. Of course, situations where parameters may be invariant to some policies and not others are common in structural estimation – see, e.g., Keane and Wolpin (2001) for a discussion of this issue in the context of their model of college attendance decisions, which can only be used to simulate policies that leave parents' decision rule for providing financial support to children unaffected.

Here, if $\gamma_5 \neq 0$, then, in interpreting estimated effects of childcare on child outcomes, we must be careful to view them as applying only to policy experiments that do not alter the decision rule for goods investment in children (6). As this decision rule is conditional on work and childcare usage decisions, it will be invariant to policies that leave the budget constraint conditional on those decisions unchanged. A work requirement that induces a woman to work and use childcare, but that leaves her wage rate and the cost of care unaffected, would fall into this category.

In contrast, a policy like childcare subsidies would shift the budget constraint conditional work and childcare usage, so it is unlikely to leave (6) unchanged. Such subsidies would not only alter childcare use, but potentially also goods inputs, and in a way not captured by $\alpha_3 \cdot \gamma_5$.

To our knowledge, the issues we are discussing here were first raised by Rosenzweig and Schultz (1983). They call an equation like (7), where proxy variables are substituted for one or more unobserved inputs, a "hybrid" production function, and they discuss the potential problems that may arise in interpreting estimates of such a function. Rosenzweig and Wolpin (1994) and Todd and Wolpin (2003, 2007) also discuss this issue. It is important to note, however, that there is no ideal way out of this problem. The alternative to proxying for unobserved goods inputs is to simply ignore them, which would lead to omitted variable bias. As noted by Todd and Wolpin (2007), it is not obvious *a priori* which approach would lead to greater bias.

Appendix C: Tests of the Pooling Hypothesis

It is desirable to pool the PPVT, PIAT-M and PIAT-R tests for use in estimation of equation (9) in the text, as this leads to a substantial efficiency gain (i.e., due to increased sample size). However, before pooling the data it is important to test that the production function parameters are invariant across the three tests. We consider two types of test of the pooling hypothesis:

The first type of test is based on differences in scores on two different tests for the same child at the same age. Let S_{it}^1 and S_{it}^2 denote scores of child *i* on two different tests both measured at the same age *t* (e.g., test 1 could be the PPVT and test 2 could be the PIAT-M, both recorded at age 5). Let W_{it} denote the complete set of variables included in equation (9). Then, if we run the regression:

(C1) $$\ln S_{it}^1 - \ln S_{it}^2 = W_{it}\Gamma + \varepsilon_{it},$$

under the null hypothesis that all parameters of (9) are equal for each test we should not be able to reject H_0: $\Gamma=0$. More generally, let $W_{it} = (U_{it}, X_{it})$ where U_{it}, is a subset of variables whose effects are invariant across tests, and X_{it} is a subset of variables whose effects differ by test. Then if we run the regression:

(C2) $$\ln S_{it}^1 - \ln S_{it}^2 = U_{it}\Delta + X_{it}\Upsilon + \varepsilon_{it},$$

we should not be able to reject the null hypothesis that $\Delta=0$.

A second type of test involves pooling the data on all three tests together and estimating a fully interacted specification of the form:

(C3) $$\ln S_{it}^j = W_{it}\Omega + W_{it}d_{it} \cdot \Gamma + \varepsilon_{it}$$

where $W_{it}d_{it}$ is a complete set of interaction terms between W_{it} and a set of test indicators *d*, with associated coefficient vector Γ. Under the pooling hypothesis we should not be able to reject the null hypothesis that $\Gamma = 0$. More generally, suppose only the subset of variables X_{it} have effects that differ by test. Then, if we run the regression:

(C4) $$\ln S_{it}^j = U_{it}\Psi + U_{it}d_{it} \cdot \Delta + X_{it}\Phi + X_{it}d_{it} \cdot \Upsilon + \varepsilon_{it},$$

we should not be able to reject the null hypothesis that $\Delta=0$. That is, once we allow to subset of variables X_{it} to have effects that differ by test, we should not be able to reject the hypothesis that the remaining variables U_{it} have common effects.

Results of these tests are reported in Table C1 Panel A. The first four rows of panel A report results for tests of the first type – i.e., regressions of test score <u>differences</u> on all regressors in equation (9). Note that we can only examine four pairs of tests, because not all tests are observed at all ages (i.e., only the PPVT and not the PIATs are taken at ages 3 and 4). The first column contains tests of the type C1, where no conditioning variables are allowed.

For the PIAT math and reading tests at ages 5 and 6 we cannot reject the null hypothesis of common coefficients at the 5% level (p-values of .308 and .055 respectively). However, for the difference between the PPVT at age 5 and either the PIAT math or reading at age 5 we do reject the hypothesis of common coefficients (p-values of .025 and .018 respectively).

Thus, we turn to the second column of Table C1, which contains tests of the form C2. Here, we include four conditioning variables in X_{it}: These are race (i.e., non-white) interacted with two test dummies (PIAT-R is the omitted category) as well both AFQT and the AFQT missing indicator interacted with the PPVT dummy. These variables were chosen because they were significant in the PPVT-PIAT regressions. As we see in column (2), with these variables included in X_{it} we cannot reject the null hypothesis that all other coefficients are zero (i.e., Δ=0) at the 10% level.

The last row of Panel A reports tests of the second type – i.e., regressions where we fully interact all variables in equation (9) with test indicators, allowing their effects to differ freely by test. The first column reports tests of the type C3, where we test if all the interaction terms are jointly significant. The hypothesis that all interactions are zero (i.e., $\Gamma = 0$) is clearly rejected (p=.0000).

However, in the second column we report tests of the type C4, where we include in X_{it} the same four controls as before: interactions of race (i.e., non-white) with test dummies and interactions of AFQT and AFQT missing with the PPVT dummy. With these controls included in X_{it}, we can no longer reject the null hypothesis that all other test interactions are zero (i.e., Δ=0) at the 5% level.

Thus, we conclude that in our baseline model it is appropriate to pool the three tests <u>provided</u> interactions of race (i.e., non-white) with test dummies and interactions of AFQT and AFQT missing with the PPVT dummy are included in the model.

Table C2 contains a comparison of our baseline results with and without these interactions. Note that failure to include them leads to some upward bias in the cumulative childcare coefficient, from -0.53% per quarter to -0.73% per quarter.[23]

The interaction coefficients are interesting in themselves. The coefficient on the AFQT with PPVT interaction is .0022 with a t-stat of about 5.5. As the standard deviation of AFQT is 18.30, this

[23] Note that the model without interactions uses 16 factors as instruments while the model with interactions uses only 14 factors. In each case the factors were chosen using the method described in Section 6.3, and the candidate factors were identical. But to retain a factor as an instrument we required that it have a t-statistic of at least 3 in at least one of the first stage regressions for the 3 endogenous variables. Once the interaction terms were added to the first stage regressions, 2 of the factors no longer satisfied the t>3 cutoff, so the number of IVs dropped from 16 to 14. However, results obtained using all 16 instruments in both models were essentially identical.

implies that a one standard deviation increase in the mother's AFQT raises the PPVT score relative to the PIAT scores by (.0022)(18.30)=4%. The race coefficients imply that, *ceteris paribus*, whites and non-whites have essentially identical scores on the PIAT-R, but that non-whites have scores 11.5% lower on the PPVT and 1.9% lower on the PIAT-M. Why PPVT scores are more sensitive to mother's race and AFQT than PIAT scores is a very interesting question for future research, but it goes beyond the scope of the present investigation.

Table C1 Panel B reports results where we let the cumulative childcare coefficient differ by test. (Otherwise this model is exactly the same as our baseline specification in Table C2 that includes all the interaction terms noted above). We report results for LIML using the 14 factors as instruments.[24] The point estimates imply that the cumulative childcare effects are greater for the PIAT-M and PIAT-R (-0.50% and -0.61% per quarter, respectively) than they are for the PPVT (-0.32% per quarter). And indeed, while the coefficients for the PIAT tests are significant at the 5% level, that for the PPVT is not. However, the χ^2 test for the hypothesis that the three childcare coefficients are equal is 1.53 which has a p-value of .464. Thus, there is no clear evidence of differential effects of childcare by test. Based on these results, we conclude it is appropriate to pool the tests (provided the interaction terms noted above are included in the model).

Finally, we repeated the above analysis breaking the non-white indicator into separate black and Hispanic indicators. This produced almost identical results. That is, given interactions of both black and Hispanic with the test dummies (and the AFQT/PPVT interaction), we could no longer reject pooling. Having separating indicators for blacks and Hispanics increased the estimated childcare effect only slightly from -0.53% to -0.57% per quarter.[25] Also, Table 10 reports results from a model that lets childcare effects differ by race. We do not find significant differences.

Appendix D: Testing the Cumulative vs. Current Childcare Specifications

Table D1 compares estimates of specifications where current test scores depend on cumulative or current childcare. The cumulative childcare variable sums up quarters of childcare (with a quarter of full-time care counting as 1 and a quarter of part-time care counting as 0.5). In the cumulative specification in column (1) the estimated cumulative childcare coefficient is -.00533 (with a standard error of .0025). This means that each additional quarter of full-time childcare reduces test scores by roughly 0.53%. This corresponds to an effect of -2.1% per year.

[24] Results using the full set of 78 excluded instruments are quite similar.

[25] It is not surprising that separating out blacks and Hispanics has little effect on the results, because the estimated test dummies for blacks and Hispanics are fairly similar. For blacks PPVT scores are 13.8% lower than whites, *ceteris paribus*. For Hispanics the figure is 10.2%. For the PIAT-M the analogous figures are 2.8% and 2.1%. For the PIAT-R blacks and whites have essentially identical scores, while the scores for Hispanics are 2.7% lower.

In the current specification in column (2), the estimated effect of current childcare is -0.0269 (standard error .0137). This implies that a single quarter of full-time childcare reduces test scores by 2.7%. This very large effect estimate is not surprising, given the very strong persistence in childcare use. The probability a mother uses childcare in quarter t conditional on having used it in quarter t-1 is 93.5%. Similarly, the own transition rate for non-use of childcare is 89.1%.[26] Thus, current childcare use provides a very good proxy for lagged childcare use.

We can formally test the cumulative specification vs. the current specification in the following way: First, we can reject the current specification if lagged childcare inputs matter conditional on current inputs. Second, the cumulative specification implies the restriction that the current and lagged childcare coefficients are all equal. If we find that lagged inputs matter but that coefficients are not equal it implies hat a more general dynamic model is appropriate.

If we add four years of lagged childcare indicators to the current specification, the p-value for their joint significance is .0329. Thus, we reject that only current childcare matters. Furthermore, a χ^2 test for equality of coefficients on current and lagged childcare gives a p-value of .1748. This supports the cumulative specification. Admittedly, however, a cumulative specification would be hard to reject, given the great persistence in childcare use noted above. As a result of that persistence, lagged childcare indicators are highly collinear, and individual lag coefficients are very imprecisely estimated. This is why we do not report coefficients on the individual lags.

Appendix E: Results Using Principle Components as Instruments

Our estimation procedure involves factor analyzing the instruments, rotating the factors by the commonly used varimax method, and then estimating each factor using the commonly used regression method. We then run regressions of the endogenous variables on the factors, and choose the subset of factors that have the greatest explanatory power for the endogenous variables. This led us to choose 14 factors that had t-statistics of at least 3 in those regressions.

The factor rotation serves two valuable purposes in this procedure: First, it gives us orthogonal factors, thus leading to low correlations amongst our instruments. Second, it gives us factors with a clear interpretation (e.g., factor 6 measures welfare benefit levels) so we can judge if the factors have sensible signs in the first stage regressions.

An alternative procedure is to simply use principle components factor analysis, and choose factors that have eigenvalues above some cutoff level (e.g., a commonly used eigenvalue threshold for retaining a factor is 1). We could then estimate these factors, and use the estimated principle

[26] These transition rates are calculated only over the 12 quarters after birth, for which we observe childcare use directly.

factors as instruments. [Note: Factors must always be estimated in a second stage after factor analysis is performed, regardless of whether one uses rotated factors or principle components. The estimated factors are simply weighted sums of the original instruments].

In our view there are several fundamental problems with this principle factor approach. First, note that the most important principle factors (i..e, those with the largest eignevalues) are the ones that explain most of the covariance *amongst the instruments*. This is not at all the same thing as the set of factors that will best explain the *endogenous variables*.

To see the difference, consider the following simple example. Suppose we factor analyze 9 variables. The first 8 are highly correlated amongst each other, while the 9th has low correlations with the first 8. A factor analysis of these data produces two factors. The first accounts for the covariances amongst the first 8 variables. It has a very large eigenvalue because it explains most of the covariance in the data. There will also be a second factor, but it has a very small eigenvalue because only the 9th variable loads on it (i.e., it explains little of the covariance in the data). Now, this scenario is perfectly consistent with a situation where the 9th variable, and hence the 2nd factor, has far more explanatory power for the endogenous variables than does the first factor. But a researcher choosing instruments using a principle factors/largest eignevalues criterion might mistakenly conclude that only the first factor is important, and discard the relatively unimportant second factor – leading to a very inefficient IV procedure.

As the above example illustrates, factors will have large eignevalues simply because they capture correlations among a large number of variables, regardless of whether those variables have much explanatory power for the endogenous variables. Our own results give a very clear illustration of this phenomenon. Note that it is the factor with the 6th largest eigenvalue that has the most explanatory power for childcare usage. This factor captures welfare benefit levels.

In contrast, the factor with the largest eigenvalue is the one that captures time limits. Why? Simply because we use 8 variables to measure time limits, and only two variables to measure benefit levels. So of course the factor that captures time limits has a much larger eigenvalue, because it captures a much larger set of covariances among the instruments. But, as Fang and Keane (2004) note, time limits are not nearly as important as benefit levels in explaining welfare participation.

Despite these arguments, as an experiment we decided to compare our results to those obtained using unrotated principle factors as instruments. The results are reported in Table E1. First, we report results using an eigenvalue cutoff of 1.0 to retain factors. This leads us to retain 13 factors, which explain 91.8% of the covariance among the original 78 instruments. Estimates of these factors are then used as instruments. Given these instruments, the estimate of the childcare effect is -0.38% per quarter and it is not significant.

Reducing the eigenvalue cutoff to 0.5 leads us to retain 18 factors, which explain 96% of the covariance among the original 78 instruments. This gives an estimated childcare effect of -0.51% per quarter which is significant at the 10% level.

Finally, reducing the eigenvalue cutoff to 0.3 leads us to retain 21 factors, which explain 97.3% of the covariance amongst the 78 instruments. This gives an estimate of the childcare effect of -0.63% per quarter which is significant at the 5% level. It is only at this point that the estimates "settle down" in the sense that adding more factors did not noticeably alter results.

The last row of Table E1 reports our baseline results using 14 selected rotated factors. Three advantages of our procedure are apparent. Our 14 factors produce a more efficient estimate (i.e., a standard error of .0025 vs. .0027 when we use the 21 principle factors). And our 14 factors have a higher correlation with the endogenous variables than do the 21 principle factors (e.g., a Shea partial R^2 for childcare of .0967 vs. .0960). And we obtain a higher value of the Cragg-Donald weak instrument statistic (15.33 vs. 11.13). Note that the third advantage actually follows directly from the first two: the weak instrument statistic is higher because we obtain higher correlations with the endogenous variables using fewer instruments.

Appendix F: Alternative Tests for Effect of Mother's Age at Birth of Child

As we indicated in the main text, our welfare policy instruments are correlated with mothers' age at childbirth, due to the timing of waivers/TANF and the structure of the NLSY79. Specifically, a child had to be born in 1990 or later to have any chance of being affected by the reform before the age of 6. And women who had children prior to 1990 tend to be younger at childbirth than those who had children later. Indeed, from 1990 onward, all births are to mothers in their 20s and 30s, while prior to 1990, many were to teenage mothers. The youngest women in the NLSY – i.e., those who were 14 in Jan. 1979 – would be 24 by Jan. 1990. Thus, the large majority of mothers who would have been affected by welfare reform would have been at least 24 at childbirth.

So we have that welfare reform: (a) positively affected maternal work/childcare use, and (b) is positively correlated with age at childbirth. This correlation alone will not generate bias. However, if: (i) mother's age at birth has a positive effect on child cognitive ability, and (ii) we fail to adequately control for mother's age in the main equation, this will generate a spurious *positive* effect of maternal work/childcare use on child cognitive test scores. As we actually find a negative effect, the plausible concern is that we *understate* this negative effect.

Table 8 column (2) in the main text presents an alternative specification that adds additional controls for mother's age at childbirth to the main equation. Specifically, we add age and age squared (whereas the baseline model contains only dummies for whether the mother was under 20 or over 33). The inclusion of the additional controls has a negligible impact on the estimates (compare Table 8 columns (1) and (2)). In other results not reported we also tried eliminating all age controls from the main equation, and found this also has a negligible effect (the estimated childcare effect ranges from -0.49% per quarter with no controls for age at all, to -0.52% with all four controls).

Strikingly, inspection of the coefficients on the mother's age controls reveals that mother's age at childbirth has little effect on child cognitive outcomes (conditional on measures of the mother's human capital like education, AFQT). Thus, the potential source of bias noted in (i) above – i.e., a *positive* association between maternal age at birth and child cognitive outcomes – is simply not present. Indeed, the column (1) estimates imply that, *ceteris paribus*, children of teenage mothers have 2.4% *higher* test scores. Thus, if anything, it seems that (controlling for economic resources) maternal youth is slightly beneficial for child outcomes. But this effect is too weak for the presence or absence of age controls to have much impact on the childcare coefficient.

To further address the age issue, in Table F1 we report estimates for sub-samples of women based on age at childbirth. We restrict the sample to women who were 24+, 24-34 or 24-30 years old at childbirth. For these sub-samples, the estimated effects of childcare range from -2.3% to -1.6% per year. However, due to the reduced sample size these estimates are not significant. So in columns (4)-(6) we use the full sample, but we interact the childcare coefficient with dummies for whether the mother's age at childbirth was outside the indicated range. None of the interactions is significant, and the main effect of childcare is quite stable across the three columns (i.e., it ranges from -2.3% to -2.4% per year and is always significant). Thus, we find no evidence that the age/welfare policy correlation leads to significant bias in our estimates of the childcare effect.

Appendix G: Alternative Tests for State Effects

As we noted in Section 6.4, a possible concern with our results is cross-State correlation between the instruments and *unobserved* child skill endowments. That is, it is possible that States where children had relatively low unobserved skill endowments may have adopted stricter welfare reform. This would bias our estimated childcare effect negatively.

Here we test for the existence of such a correlation. Specifically, in Table G1 we group States into those that adopted more vs. less strict approaches to welfare reform along five different dimensions. Then we compare average test scores in the pre-reform period between each group of

States. In each case, there is no significant difference in average child test scores in the pre-reform period between States that subsequently adopted more vs. less strict welfare reform programs. Thus, there is no evidence of cross-State correlation between the instruments and child skill endowments.

Appendix H: Alternative Approaches to Clustering

In our main results we have clustered the standard errors by child. In our data, there are an average of 2.59 test score observations per child, and we would expect strong correlations among these scores, arising from child latent ability endowments. Indeed, if we do an analysis of variance of the residuals from our main equation, we find that child effects account for 57% of the variance of the residuals. In order to assess the importance of clustering, Table H1 presents results with and without clustering, for four different specifications of the instrument set.

Specifically, Table H1 reports estimates for our baseline specification of the main equation estimated using: (1) the full set of 78 instruments, (2) only State level instruments, but including their interactions with mother's education and AFQT, (3) only State level instruments with no interactions, and (4) the 14 instruments obtained via factor analysis.[27] Standard errors obtained without clustering are reported directly under the estimates, and standard errors clustered by child are reported below those. As we see, for the childcare coefficient, clustering by child increases standard errors by 15% to 40%, depending on the instrument set. This large increase is not surprising, given the high within-cluster correlation of the errors. It is interesting that the largest increase is in the model with all 78 instruments (40%) while the smallest is in the model that uses only the 14 factors as instruments (15%).

One might argue that we should cluster at a higher level. For instance, one could argue we should cluster at the level of the mother, as 368 of our 944 mothers have multiple children, and there may be unobserved mother effects that are common in the family. Or one could argue we should cluster at the level of the State, as much (but certainly not all) of the variation in our instruments is at the State level. These results are also reported in Table H1.

Note that clustering by State is not possible for the models in the first two columns where the number of instruments is 78 and 63, respectively. This is because the construction of the robust covariance matrix estimator requires that there be more clusters than instruments. In the last two columns the number of instruments is 25 and 14 respectively, so it is possible to cluster by State. Even so, we would view the State clustered standard errors with some caution, as the asymptotics of

[27] The specifications correspond to: Table 9 column (1), Table 9 column (7), Table 9 column (8) and Table 6 column (6), respectively.

the robust covariance matrix estimator relies on the number of clusters growing large relative to the number of moment conditions (instruments).

The results show that clustering by mother leads to slightly larger standard errors than clustering by child, while clustering by State leads to slightly smaller standard errors. However, these differences are minor compared to those between clustering by child vs. not clustering at all. Thus, the choice of clustering level makes essentially no difference to any of our results.

Still, we prefer the child-clustered results for a number of reasons: First, the number of State clusters is too small relative to the number of instruments to have great faith in the asymptotic theory (Indeed, we can only implement State clusters in a small subset of our specifications). And, in an analysis of variance of the residuals, State effects account for only 2% of the variance. Thus, they seem to be quite unimportant compared to child effects, which account for 57%.

Second, regarding clustering by mother, we note that mother effects can explain 17% of the variance of the residuals. Thus, mother effects are not as important as child effects, and the latter completely subsume the former. An issue with mother effects is that the within cluster correlation is much higher for clusters that contain only one child than for clusters that contain two or more children. Hence, the intra-cluster correlation will be a weighted average of the two, and this mongrel parameter will be too low for the former and too high for the latter. Intuitively, the larger the intra-cluster correlation, the greater is the downward adjustment of the sample size needed to obtain the effective sample size used to compute standard errors. Our intuition is that this will cause the single-child observations to be counted too much and the multi-child observations counted too little.

Appendix I: Robustness Checks for Informal vs. Formal Childcare Results

As we noted in Section 6.7, our finding that informal childcare has adverse affects relative to formal care is arguably the most important of this study – in that it may provide a rationale for government programs (like CCDF in the U.S. or Child Care Benefit in Australia) that create incentives for mothers to use formal rather than informal care. Thus, we subjected this result to the same battery of robustness tests we applied to our estimates of the effect of childcare in general. Before we begin, recall that we found (see Table 11) that a year of informal care reduces test scores by 0.64% per quarter or 2.6% per year (t=-2.22) while formal care has a positive but insignificant effect (+0.30% per quarter or +1.2% per year, t=0.46).

In Table I1 we report the same battery of sensitivity tests to changes in specification of the main equation that we reported in Table 8. First consider robustness to controls for mother's age at childbirth. As we see in column (2), adding additional age controls only shifts the informal childcare coefficient slightly, from -0.64% to -0.68% per quarter, and it remains significant at the 5% level. Again, the age coefficients show little effect of mother's age at childbirth on child outcomes.

In Table I2 we report results analogous to Table F1, where we restrict the sample to mothers who were 24+, 24-34 or 24-30 years old at childbirth. Here the loss of sample size causes the estimates to lose significance, but the point estimates in the -0.47% to -0.72% range are still broadly consistent with our earlier results.[28]

We now return to Table I1 and consider sensitivity of our results to five other changes in the specification of the main equation. First, we drop the mother's AFQT score. This increases the estimated informal childcare coefficient from -2.6% to -3.5% per year. As in Table 8, it also produces a large increase in the cumulative income coefficient, which becomes highly significant. As we noted in Section 6.4, this suggests that, with AFQT omitted, transitory income becomes more important, as it proxies for the mother's permanent income/skill endowment.

In the next column we consider sensitivity of our results to controls for the ages of siblings. Including separate regressors for number of children aged 0-5 and 6-17 (both treated as endogenous) has essentially no impact on the estimated informal and formal childcare effects.

Next, we consider aggregate time effects. As in Table 8, when we include a quadratic time trend it has a U-shape, implying an aggregate factor not included in our model first drove down test scores followed by a recovery. But including the time trend only slightly reduces the estimated informal childcare effect, from -2.6% to -2.4% per year, and the formal childcare effect is little changed. Thus, any bias from omitted time effects appears to be minor. Results were essentially identical using unrestricted time dummies.

In the next column of Table I1 we add State fixed effects to the main equation. This increases the estimated cumulative informal childcare effect from -.64% per quarter to a very large value of -1.01%. But this is imprecisely estimated and hence only significant at the 10% level. This large increase in the informal childcare coefficient is similar to what happened in Table 8 for all childcare.

However, as we pointed out in Section 6.4, we are skeptical of the fixed effects estimates for several reasons. First, as we already noted, State effects account for only 2% of the variance of the residuals. Second, as we see in Table I1 column (6), State effects are not jointly significant at the 10% level, so there is no clear evidence for including them. Third, even if unobserved child ability did differ by State, it would only induce bias if it were correlated with the instruments – e.g., if

[28] In Table F1 we deal with this small sample problem by also using the full sample and interacting the childcare variable with indicators for mother's age at childbirth. The same approach was not successful here because now we have two interactions (age at childbirth with formal and informal care) and thus two new endogenous variables. We were not able to find reasonably powerful instruments for the interaction between formal care and mother's age at childbirth.

States with low ability children adopted stricter reforms. This strikes us as implausible a priori, and Web Appendix G presents empirical evidence that is contrary to this idea. Fourth, in the childcare context we are skeptical of the strict exogeneity assumption required for consistency of the fixed effects estimator.[29] Finally, Keane and Wolpin (2002) show that fixed effects can lead to very misleading results if expected future values of policy variables matter for current decisions.[30]

However, our fixed effects results are at least comforting in that they suggest that failure to include fixed effects does not bias our estimates of the informal childcare effects in a negative direction.[31] That is, the childcare coefficient becomes even <u>more</u> negative when State effects are included. So there is no indication that failure to account for State fixed effects is driving our result that informal care has a negative effect.

Finally, in the last column of Table I1 we consider a model where test scores enter in levels. The estimates imply that a quarter of informal childcare reduces scores by 0.49 points. As the mean score is 91.9, this is -0.53% per quarter, or -2.1% per year. This is a bit smaller then the -2.6% effect estimated in logs. The estimate for formal care is +0.40% per quarter, but still insignificant.

In Table I3 we consider sensitivity of our formal/informal childcare results to the specification of the instrument list. This table is analogous to Table 9 in the main text. It reports LIML results using the baseline list of 78 instruments in column (1), and seven variants on that list in columns (2)-(8). As in the main text, we must use the full set of 78 instruments rather than the 14 factors for this exercise because we want to consider the impact of dropping certain types of variables from the instrument list (and there is not a one-to-one mapping between factors and types of instruments, because some factors pick up multiple aspects of welfare reform).

In column (2) we exclude CCDF spending. Excluding this instrument increases the estimated informal childcare effect slightly from -2.6% to -2.8%. In column (3) we use only the main features of TANF as instruments: time limits, work requirements and disregards. This increases the estimated informal childcare effect substantially to –3.8% per year.

---

[29] The strict exogeneity assumption will fail if children's test score realizations at age *t* affect future inputs into child production, and/or how the welfare policy rules evolve. [See also the discussion of this point in Section 2].

[30] A State fixed effect controls for a State's average level of welfare generosity. Thus, using State effects, we estimate the impact of <u>deviations</u> from the average level of childcare use induced by <u>deviations</u> from average welfare rules. Such short-run effects may differ from effects of long-run policy changes, and the latter are presumably of greater interest.

[31] A negative bias would emerge if States where children had low skill endowments adopted stricter welfare reform (making it look like stricter welfare rules led to more work/childcare which in turn led to lower scores). Recall that we looked at this issue directly in Web Appendix G. There we group States into those that adopted more vs. less strict approaches to welfare reform along five different dimensions. Then we compare average test scores in the pre-reform period between each group of States. In each case, there is no significant difference in average child test scores in the pre-reform period between States that subsequently adopted more vs. less strict welfare reform programs. Thus, there is no evidence of cross-State correlation between the instruments and child skill endowments.

In column (4) we drop all the TANF related instruments, using other aspects of policy and demand environment to identify the childcare effect. In column (5) we go even further and drop all instruments specific to the welfare reforms of the 90s (e.g., TANF, CCDF, EITC), using only instruments that would have varied across States/time regardless. These are State welfare grant levels and local demand conditions (i.e., State unemployment rates and 20th percentile wage rates).

Clearly the attempts to apply IV in columns (4) and (5) were not successful. These smaller instrument sets do not contain variables that are effective at predicting whether people use formal vs. informal childcare. Recall that in Section 6.7 we noted that mothers are more likely to use formal relative to informal care if: (i) a State does *not* have a work requirement, (ii) it has young child or other work requirement exemptions, (iii) it has a longer work requirement time limit, (iv) work requirements were implemented more recently, (v) less time has elapsed since a time limit could have hit, (vi) remaining eligibility is greater, (vii) a State has higher CCDF spending, or (viii) earnings disregards are greater. Thus, all of the effective predictor variables are aspects of the TANF reforms or waivers. When we drop all of these variables, the point estimates for cumulative formal and informal childcare care become very imprecise. And, in columns (4) and (5), the Cragg-Donald weak instrument test statistic falls to very low values of 0.81 and 0.44, respectively. Thus, we view the results in columns (4) and (5) as uninterpretable.

Recall that in our reduced form regression, we interact all policy and demand variables with mother's education and AFQT. This allows changes in welfare policy and demand conditions to have different effects on different types of mothers (e.g., welfare rules are less important for the college educated). In column (6) we drop these interactions to see how important they are. The estimated informal childcare effect is reduced only slightly to -2.5% per year vs. -2.6% under the baseline. The formal childcare estimate is hardly changed.

Next, recall that some of our instruments are tailored to individuals based on ages of their children (e.g., whether a woman could have reached the time limit – see Section 3.1). In column (7) we drop these individual specific instruments, and rely purely on cross-State and over time variation to identify the childcare effect. The resulting estimate of the informal childcare effect is -3.1% per year, which is a bit larger than our baseline estimate. In column (8) we go further and also drop the interactions of the remaining instruments with mother's education and AFQT. This gives an estimate of -2.9% per year.

In summary, our result of a negative effect of informal childcare is robust to a wide range of alternative instrument sets, with estimates ranging from –2.5% to –3.8% per year, and all but one estimate between -2.5% and -3.1% (compared to our baseline of -2.6%). Our finding of no adverse effect of formal childcare is also robust across all these instrument sets.

The only exception to this pattern was when we dropped instruments related to TANF and other aspects of the welfare reforms of the 1990s. These instruments are crucial for predicting whether mothers use formal or informal care. This is because, loosely speaking, reforms that had a "carrot" like aspect (i.e., that created positive incentives to work) tended to induce more use of formal care, while reforms that took more of a "stick" approach (e.g., work requirements) tended to induce more use of informal care. Thus, without these instruments, we are unable to identify separate informal and formal childcare effects.

Appendix J: Who Uses Formal vs. Informal Care, and Relatives vs. Non-Relatives?

In this section we present evidence on what type of mothers are more likely to use different types of childcare. In the first column of Table J1 we present a logit for whether a mother uses formal or informal childcare (conditional on childcare use). The results show that more educated, urban women with fewer children are more likely to use formal care. This suggests that formal care is higher quality, as it is typically used by women who can afford more expensive care.

In column 2 of Table J1 we present a logit for use of relatives vs. non-relatives (conditional on using informal care). It is the more educated, urban women with fewer children who are more likely to use non-relatives. Again, this is suggestive that non-relatives provide higher quality care than relatives.

These findings about who uses each type of care appear consistent with our finding that only informal care has adverse effects on child cognitive development, and that, amongst types of informal care, only care by relatives has a significant negative effect.

Web Appendix References

Akerman-Ross, S. and P. Khanna, 1989, “The Relationship between High Quality Day Care to Middle-class 3-year-olds’ Language Performance,” *Early Childhood Research Quarterly* 4:97-116.

Baydar, N. and J. Brooks-Gunn, 1991, "Effects of Maternal Employment and Child Care Arrangements on Preschoolers' cognitive and Behavioral Outcomes: Evidence from Children of the National Longitudinal Survey of Youth," *Developmental Psychology*, 27(6), 932-945.

Ben-Porath, Y., 1967, “The production of human capital and the life-cycle of earnings,” *Journal of Political Economy*, vol. 75, no. 4, part I, August, 352-265.

Bernal, R., 2008, “The Effect of Maternal Employment and Child Care on Children's Cognitive Development,” *International Economic Review*, Vol. 49, No. 4, pp. 1173-1209.

Blau, D., 1999a, “The Effects of Child Care Characteristics on Child Development,” *The Journal of Human Resources*, XXXIV, 4.

Blau, F. and A. Grossberg, 1992, "Maternal Labor Supply and Children's Cognitive Development," *The Review of Economics and Statistics*, 74(3), 474-481.

Blau, D. and J. Currie, 2004, “Preschool, Day Care, and Afterschool Care: Who’s Minding the Kids” in Handbook of the Economics of Education, Hanushek, E. and F. Welch eds. Elsevier.

Burchinal, M.R., S. Ramey, M. Reid and J. Jaccard, 1995, “Early Child Care Experiences and their Association with Family and Child Characteristics during Middle Childhood,” *Early Childhood Research Quarterly*, 10: 33-61.

Caughy, M., J. DiPietro and D. Strobino, 1994, “Daycare Participation as a Protective Factor in the Cognitive Development of Low-Income Children,” *Child Development*, 65:457-471.

Chase-Lansdale, P.L., R. Moffitt, B. Lohman, A. Cherlin, R. Coley, L. Pittman, J. Roff, E. Votruba, 2003, “Mothers’ Transitions from Welfare to Work and the Well-Being of Preschoolers and Adolescents,” *Science*, 299(March):1548-1552.

Clarke-Stewart, K., C. Gruber, L. Fitzgerald, 1994, *Children at Home and in Day Care*, Hillsdale.

Currie, J. and D. Thomas, 1995, “Does Head Start Make a Difference?,” *American Economic Review*, 85:3, 341-364.

Currie, J. and D. Thomas, 2001, “Early Test Scores, School Quality and SES: Long Run Effects on Wage and Employment Outcomes,” Worker Wellbeing in a Changing Labor Market, Vol. 20.

Desai, S., L. Chase-Lansdale and M. Robert, 1989, "Mother or Market? Effects of Maternal Employment on the Intellectual Ability of 4-year Old Children," *Demography*, Vol. 26, No. 4.

Duncan, G. and NICHD Early Child Care Research Network, 2003, “Modeling the Impacts of Child Care Quality on Children’s Preschool Cognitive Development,” manuscript Northwestern University.

Dunn, L., 1993, “Proximal and Distal Features of Daycare Quality and Children’s Development,” *Early Childhood Research Quarterly* 8:167-192.

Fang, H. and M. Keane, 2004, “Assessing the Impact of Welfare Reform on Single Mothers,” *Brookings Papers on Economic Activity*, Vol. 1.

Greenstein, T., 1995, "Are the 'Most Disadvantaged' Children Truly Disadvantaged by Early Maternal Employment?," *Journal of Family Issues* 16(2), Mar: 149-169.

Han, W., Waldfogel, J., and J. Brooks-Gunn, 2001, "The Effects of Early Maternal Employment on Later Cognitive and Behavioral Outcomes," *Journal of Marriage and the Family*, 63(2): 336-354.

Harmon, C. and I. Walker, 1998, "Selective Schooling and the Returns to the Quantity and Quality of Education," manuscript Department of Economics, University College Dublin.

Harvey, E., 1999, "Short-term and Long-term Effects of Early Parental Employment on Children of the National Longitudinal Survey of Youth," *Developmental Psychology* 35(2), 445-459.

Haveman, R. and B. Wolfe, 1994, Succeeding Generations: On the Effects of Investments in Children. New York: Russell Sage Foundation.

Howes, C. and J. Rubenstein, 1981, "Toddler Peer Behavior in Two Types of Day Care," *Infant Behavior and Development*, 4:387-393.

______. 1985, "Determinants of Toddlers' Experiences in Day Care: Age of Entry and Quality of Setting," *Child Care Quarterly*, 14: 140-151.

Hutchinson, D., H. Prosser and P. Wedge, 1979, "The Prediciton of Educational Failure," *Educational Studies*, 5(1).

James-Burdumy, S., 2005, "The Effect of Maternal Labor Force Participation on Child Development," *Journal of Labor Economics*, Vol. 23(1): 177-211.

Keane, M.P. and K. Wolpin, 2001, "The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment," *International Economic Review*, vol. 42(4), pages 1051-1103, November.

Keane, M. and K. Wolpin, 2002, "Estimating Welfare Effects Consistent with Forward Looking Behavior, Lessons from a Simulation Exercise," *Journal of Human Resources*, 37:3, 570-622.

Lamb, M., 1996, "Effects of Nonparental Child Care on Child Development: An Update," *The Canadian Journal of Psychiatry*, 41: 330-342.

Leibowitz, A., 1974, "Home investments in children," *Journal of Political Economy*, 82:2, Part I, 111-131.

Liu, H., T. Mroz and W. van der Klaauw, 2003, "Maternal Employment, Migration, and Child Development", manuscript University of North Carolina at Chapel Hill.

Love, J.M., P. Schochet and A. Meckstroth, 1996, "Are They in Real Danger? What Research Does -And Does not- Tell Us About Child Care Quality and Children's Well-being", Mathematica No. ED145 030.

Kristin A. Moore, K. and A. Driscoll, 1997, "Low-wage maternal employment and outcomes for children: a study," *The Future of Children*, Vol. 7, No. 1, 122-127. Spring, 1997.

Mott, F., 1991, "Developmental Effects of Infant Care: The Mediating Role of Gender and Health," *Journal of Social Issues* 47(2):139-158.

Murnane, R. J., J. Willet and F. Levy, 1995, "The growing importance of cognitive skills in wage determination," *Review of Economics and Statistics*, LXXVII, pp. 251-266.

Neal, D. and W. Johnson, 1996, “The Role of Premarket Factors in Black/White Wage Differences,” *Journal or Political Economy*, 104(5), pgs. 869-895.

NICHD Early Child Care Research Network, 2000, “The Relation of Child Care to Cognitive and Language Development,” *Child Development*, 71(4): 958-978.

Parcel, T. and E. Menaghan, 1990, "Maternal Working Conditions and Children's Verbal Facility: Studying the Intergenerational Transmission of Inequality from Mothers to Young Children," *Social Psychology Quarterly*, 53: 132-147.

Parcel, T. and E. Menaghan, 1994, "Early Parental Work, Family Social Capital, and Early Childhood Outcomes," *American Journal of Sociology*, 99(4), 972-1009.

Peterson, C. and R. Peterson, 1986, “Parent-child Interaction and Day Care: Does Quality of Day Care Matter?,” *Journal of Applied Developmental Psychology*, 7, 1-15.

Peisner-Feinberg, E., M. Burchinal, R. Clifford, M. Culkin, C. Howes, S. Kagan and N. Yazejian, 2001, “The Relation of Preschool Child-Care Quality to Children’s Cognitive and Social Developmental Trajectories through Second Grade,” *Child Development* 72(5), 1534-1553.

Robertson, D. and J. Symons, 1996, “Do Peer Groups Matter? Peer Group Versus Schooling Effects on Academic Attainment?,” Centre for Economic Performance, London School of Economics, WP No. 311.

Rosenzweig, M. and P. Schultz, 1983, “Estimating a Household Production Function: Heterogeneity, the Demand for Health Inputs, and Their Effects on Birth Weight,” *The Journal of Political Economy*, Vol. 91:5, 723-746.

Rosenzweig, M. and K. Wolpin, 1994, “Are there Increasing Returns to the Inter-generational production of Human Capital?,” *Journal of Human Resources*, 29:2, 670-693.

Ruhm, C., 2002, "Parental Employment and Child Cognitive Development", NBER WP No. 7666.

Ruopp, R., J. Travers, F. Glantz and C. Coelen, 1979, *Children at the Center: Final Results of the National Day Care Study*. Cambridge, MA: Abt Books.

Studer, M., 1992, "Quality of Center Care and Preschool Cognitive Outcomes: Differences by Family Income," *Sociological Studies of Child Development*, Vol. 5, P. Adler and P. Adler, eds.

Todd, P. and K. Wolpin, 2003, "On the Specification and Estimation of the Production Function for Cognitive Achievement," *Economic Journal*, Vol. 113(485): 3-33.

Todd, P. and K. Wolpin, 2007, “The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps,” *Journal of Human Capital*, 1(1), 91-136.

Vandell, D. and J. Ramanan, 1992, "Effects of Early and Recent Maternal Employment on Children from Low-income Families," *Child Development*, 63(4), 938-949.

Vandell, D. and M. Corasaniti, 1990, “Variations in Early Child Care: Do They Predict Subsequent Social, Emotional and Cognitive Differences?, *Early Childhood Research Quarterly*, 5:555-572.

Waldfogel, J., W. Han and J. Brooks-Gunn, 2002, "The Effects of Early Maternal Employment on Child Cognitive Development," *Demography*, May 39(2): 369-392.

Zax, J. and D. Rees, 1998, “Environment, Ability, Effort and Earnings”, Center for Research on Economic and Social Policy, Denver, Working Paper No. 9801.

Web Appendix A: Table A1

THE EFFECT OF MATERNAL EMPLOYMENT ON CHILDREN'S COGNITIVE ABILITY

(Studies that use NLSY data)

Author, year	Sample	Method	Effect of mother's employment
Mott, 1991	2387 1-4 yr olds	OLS	Negative effects
Harvey, 1999	3-12 yr olds	OLS	Negative effects
Ruhm, 2002	3-6 yr olds	OLS	Negative effects
Han et al., 2001	462 birth-8 yrs	OLS	Negative effects
Bernal, 2005	529 3 to 7 yr olds	Structural model	Negative effects
Liu, Mroz & Van der Klaauw, 2003	5 to 15 yr olds	Structural model	Negative effects
Vandel & Ramanan, 1992	1889 2nd graders	OLS	Positive effects
Parcel & Menaghan, 1994	768 3-6 yr olds	OLS	Positive effects
Greenstein, 1995	2040 4-6 yr olds	OLS	Insignificant effects
Moore & Driscoll, 1997	1154 5-14 yr olds	OLS	Insignificant effects
James-Burdumy, 2005	498 3-4 yr olds	FE and IV-FE[1]	Differing depending on test used
Waldfogel, et al., 2002	1872 birth-8 yrs	OLS and FE	Differing depending on group
Desai, et al., 1989	503 4 yr olds	OLS	Differing depending on group
Baydar & Brooks-Gunn, 1991	572 4 yr olds	OLS	Differing depending on timing
Blau & Grossberg, 1992	8784 3-4 yr olds	OLS and IV[2]	Differing depending on timing
Todd and Wolpin, 2004	6-13 yr olds	IV Child FE	Not reported

[1] Household FE, and instruments are local market conditions, e.g., county unemployment rate and percentage of the labor force in the services sector

[2] Work experience prior to childbirth is the instrument for maternal employment.

Web Appendix A: Table A2

THE EFFECT OF CHILD CARE ON CHILDREN'S COGNITIVE ABILITY

Author, year	Sample	Method	Effect of child care use
Baydar and Brooks-Gunn, 1991	572 4 yr olds	OLS	Negative effects (vary with timing)
Desai, et al., 1989	503 4 yr olds	OLS	Negative effects (only for boys)
Vandell & Corasaniti, 1990	236 8-year olds	OLS	Negative effects
Thornburg et al., 1990	835 kindergarten children	OLS	Insignificant effects
Ackerman-Ross and Khanna, 1989	3-yr olds, whites	OLS	Insignificant effects
Parcel and Menaghan, 1990	697 3-6 yr olds	OLS	Insignificant effects
Studer, 1992	95 children	OLS	Insignificant effects
Burchinal et al., 1995	6-12 yr olds	OLS	Insignificant effects
Blau, 1999	2000+ 3-5 yr olds	OLS and FE[1]	Differing depending on quality
Caughy, et al., 1994	867 5-6 year olds	OLS	Differing depending on background
Dunn, 1993	4-yr olds, middle-class	OLS	Differing depending on quality of daycare
Clarke-Stewart et al., 1994	2-4 yr olds, middle class	OLS	Differing depending on quality of daycare
Ruopp, et al., 1979	1600 preschool children	Experiment[2]	Differing depending on measure of quality
Duncan and Currie, 1995	3477 4+ yr olds	Siblings FE	Positive effects of Head Start
Peisner-Feinberg et al., 2001	773 4 - 8 yr olds	OLS	Positive effects (of high quality daycare)
NICHD Early Child Care Research Network, 2000	595 0-3 yr olds	OLS	Positive effects of center-based arrangements
Duncan-NICHD, 2003	1162 24-54 months old	OLS and FE[3]	Positive effects (of high quality daycare)

[1] Household fixed effects.

[2] The National Day Care Study randomly assigned children to classrooms with different staff-child ratios and to teachers with different levels of training. However, the 64 day care centers were not randomly selected.

[3] Child fixed effects.

Web Appendix C: Table C1
Tests of the Pooling Hypothesis

A. Significance of Test Differences

	P-value for F-test of joint significance of all variables except:	
Dependent Variable->	None	AFQT and RACE[&]
Test difference (at same age):		
PPVT5-MATH5	0.0251	0.1082
PPVT5-READ5	0.0181	0.1793
MATH5-READ5	0.3083	*
MATH6-READ6	0.0553	*
All tests pooled[**]	0.0000	0.0537

[&] AFQT*dPPVT, afqtmiss*dPPVT, RACE*dPPVT and RACE*dMATH

[#] AFQT*dPPVT, afqtmiss*dPPVT, BLACK*dPPVT, BLACK*dMATH and HISP*dPPVT.

[**] P-value for F-test of joint significance of all explanatory variables interacted with test dummies (except for what is stated in the column heading).

*Entire set of control variables insignificant, so controls for AFQT and RACE not necessary

B. Child care Effects by Test

Estimation Method	Baseline includes AFQT, NON-WHITE interacted with test dummies
LIML with 14 factors	
Child care *dPPVT	-0.00318 (0.00361)
Child care *dMATH	-0.00495 (0.00241)**
Child care *dREAD	-0.00611 (0.00258)**
χ^2 test for identical effects by test	1.53 (0.464)

** Significant at 5%; * Significant at 10%

Web Appendix C: Table C2
Tests of the Pooling Hypothesis
Dependent Variable -> Log(Test Score)

	Baseline excludes interactions of AFQT, NON-WHITE with test dummies[a]		Baseline includes AFQT, NON-WHITE interacted with test dummies[b]	
Cumulative Child Care	-0.00727		-0.00533	
	(0.0026)	**	(0.0025)	**
Log(Cumulative Income)	0.02809		0.01062	
	(0.0246)		(0.0266)	
Number of Children	-0.03036		-0.02545	
	(0.0068)	**	(0.0064)	**
I[mother's age<20]	0.02225		0.02368	
	(0.0119)	*	(0.0116)	**
I[mother's age>30]	0.00071		0.00603	
	(0.0260)		(0.0256)	
Mother's education	0.01282		0.01297	
	(0.0031)	**	(0.0030)	**
Mother's AFQT score	0.00129		0.00066	
	(0.0003)	**	(0.0003)	**
AFQT score * dPPVT			0.00220	
			(0.0004)	**
I[AFQT missing]	0.05796		0.02919	
	(0.0194)	**	(0.0181)	
I[AFQT missing] * dPPVT			0.09809	
			(0.0358)	**
I[Non-white]	-0.03999		0.00451	
	(0.0111)	**	(0.0113)	
I[Non-white] * dPPVT			-0.11986	
			(0.0208)	**
I[Non-white] * dMATH			-0.02376	
			(0.0102)	**
Male	-0.02566		-0.02593	
	(0.0071)	**	(0.0069)	**
Birthweight	0.00448		0.00555	
	(0.0063)		(0.0063)	
I[Work before]	0.02998		0.02674	
	(0.0125)	**	(0.0125)	**
I[Work before] * skill	0.00894		0.00988	
	(0.0118)		(0.0118)	
Experience before childbirth	0.00616		0.00565	
	(0.0066)		(0.0067)	
Experience * mother's age	-0.00014		-0.00013	
	(0.0002)		(0.0002)	
I[never married]	0.01807		0.01696	
	(0.0332)		(0.0318)	
I[separated]	0.03902		0.03414	
	(0.0340)		(0.0326)	
I[divorced]	0.04024		0.03218	
	(0.0355)		(0.0341)	
I[urban]	0.02188		0.02450	
	(0.0116)	*	(0.0114)	**
Child's age	0.03554		0.03817	
	(0.0113)	**	(0.0122)	**
dPPVT	-0.25327		-0.19561	
	(0.0103)	**	(0.0249)	**
dMATH	-0.07792		-0.05796	
	(0.0040)	**	(0.0092)	**
Estimation Method	LIML		LIML	
Number of Observations	3787		3787	
R-squared	0.3507		0.3844	
k [#]	1.009		1.005	
Weak/Many Instruments Test	16.45		15.33	

[a] Instruments are 16 factors derived from the factor analysis of our original 78 instruments described in the footnote in Table 5 in the main text.
[b] Instruments are 14 factors derived from the factor analysis of our original 78 instruments.
Robust standard errors (Huber-White) by child clusters.
[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

Web Appendix D: Table D1
Cumulative vs. Current Specification

Dependent Variable -> Log(Test Score)

	LIML		OLS	
	Cumulative Specification	Current Specification	Cumulative Specification	Current Specification
Cumulative Child Care	-0.00533		0.00098	
	(0.0025) **		(0.0008)	
Current Child Care		-0.02690		0.00395
		(0.0137) **		(0.0027)
Log(Cumulative Income)	0.01062	0.01503	-0.00324	-0.00311
	(0.0266)	(0.0275)	(0.0057)	(0.0056)
Mother's education	0.01297	0.01346	0.01051	0.01047
	(0.0030) **	(0.0032) **	(0.0026) **	(0.0026) **
Mother's AFQT	0.00066	0.00059	0.00059	0.00060
	(0.0003) **	(0.0003) *	(0.0002) **	(0.0002) **
Number of Observations	3,787	3,787	3,787	3,787
R-squared	0.3844	0.3682	0.3994	0.3995
Root MSE	0.1729	0.1752	0.1714	0.1714
k #	1.005	1.005		
Weak/Many Instruments Test	15.33	13.29		

Instruments are 14 factors derived from the factor analysis of our original 78 instruments described in the footnote in Table 5 in main text.
Robust standard errors (Huber-White) by child clusters.

k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

** Significant at 5%; * Significant at 10%

Web Appendix E: Table E1
Sensitivity to Factors included as Instruments

Factors	Effect of child care on log test scores (baseline specification)	Eigenvalue cutoff	Variance explained	Shea's partial R^2	Cragg-Donnald statistic
Number of principal factors:					
13	-0.00375 (0.0028)	1	0.918	0.0831	15.02
18	-0.00513 (0.0026)*	0.5	0.960	0.0925	12.27
21	-0.00633 (0.0027)**	0.3	0.973	0.096	11.13
14 selected rotated factors	-0.00533 (0.0025)**	-	-	0.0967	15.33

Web Appendix F: Table F1

Alternative Tests for Effect of Mother's Age at Childbirth

Dependent Variable -> Log(Test Score)

	(Mother's) Age Restricted Samples			Models with Age Interactions		
	24+	24-34	24-30	24+	24-34	24-30
Cumulative Child Care	-0.00585	-0.00515	-0.00404	-0.00575	-0.00565	-0.00608
	(0.0050)	(0.0045)	(0.0058)	(0.0029) **	(0.0028) **	(0.0029) **
Cumulative Child Care * Mother's age[&]				0.00153	0.00100	0.00126
				(0.0018)	(0.0011)	(0.0406)
χ2-test for additional age interactions (P-value)				0.1282	0.1040	0.1327
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	1,680	1,643	1,345	3,787	3,787	3,787
R-squared	0.3758	0.3780	0.3915	0.3872	0.3857	0.3854
k [#]	1.082	1.076	1.098	1.040	1.040	1.039
Weak/Many Instruments Test	4.25	4.49	3.85	5.68	5.72	5.72

Instruments are: All 78 policy variables, local demand conditions and interactions described in footnote in Table 5.

Robust standard errors (Huber-White) by child clusters.

[&] Mother's age dummy defined as the complement of the age stated in the corresponding column heading.

[#] k is the parameter of the k-class estimator.

** Significant at 5%, * Significant at 10%

Web Appendix G: Table G1
Average Test Scores for Children born prior to 1990

	Average	St. Dev	ttest
States that implemented TL waivers	93.34	(1.82)	-0.46
States that did not implement TL waivers	92.42	(1.08)	
States that implemented WR waivers	89.77	(1.35)	1.56
States that did not implement WR waivers	93.45	(1.09)	
States with TL lower than 3 years	90.2	(2.46)	0.87
States with TL higher than 3 years	93.02	(1.00)	
States with immediate WRs	93.48	(1.81)	-0.66
States with WRs of at least 1 month	92.20	(0.95)	
States with Age of Youngest child exemption < 6 months	93.40	(2.20)	-0.51
States with Age of Youngest child exemption > 6 months	92.38	(0.84)	

Source: NLSY, sample of single mothers

Web Appendix H: Table H1
Comparison of factor-baseline specification by type of clustering

Dependent Variable -> Log(Test Score)

	All 78 instruments		Only State-specific instruments [a]		Only State-specific IVs without interactions		14 factors as instruments	
Cumulative Child Care	-0.00522		-0.00623		-0.00642		-0.00533	
No cluster	(0.0020)	**	(0.0020)	**	(0.0020)	**	(0.0022)	**
Child clusters	(0.0028)	*	(0.0025)	**	(0.0026)	**	(0.0025)	**
Mother clusters	(0.0030)	*	(0.0027)	**	(0.0028)	**	(0.0027)	**
State clusters	-		-		(0.0024)	**	(0.0024)	**
Log(Cumulative Income)	0.01037		-0.00300		0.00529		0.01062	
No cluster	(0.0160)		(0.0163)		(0.0219)		(0.0192)	
Child clusters	(0.0242)		(0.0246)		(0.0361)		(0.0266)	
Mother clusters	(0.0255)		(0.0256)		(0.0379)		(0.0284)	
State clusters	-		-		(0.0358)		(0.0314)	
Mother's education	0.01276		0.01316		0.01167		0.01298	
No cluster	(0.0023)	**	(0.0025)	**	(0.0027)	**	(0.0025)	**
Child clusters	(0.0032)	**	(0.0034)	**	(0.0037)	**	(0.0033)	**
Mother clusters	(0.0034)	**	(0.0038)	**	(0.0040)	**	(0.0033)	**
State clusters	-		-		(0.0041)	**	(0.0034)	**
Mother's AFQT	0.00066		0.00078		0.00073		0.00077	
No cluster	(0.0003)	**	(0.0003)	**	(0.0003)	**	(0.0003)	**
Child clusters	(0.0003)	**	(0.0003)	**	(0.0004)	*	(0.0003)	**
Mother clusters	(0.0003)	**	(0.0003)	**	(0.0004)	*	(0.0003)	**
State clusters	-		-		(0.0003)	**	(0.0003)	**

[a] Excludes all individual-specific welfare rules, such as, whether a woman could have a hit a time limit or a work requirement (i.e., all variables with an i subscript in Table 1).
Standard errors in parenthesis.
** Significant at 5%; * Significant at 10%

Web Appendix I: Table I1

Robustness with respect to the Specification of the Main Equation

Dependent Variable -> Log(Test Score)

	Baseline	Aditional mother's age controls	Removing AFQT	Children by ages	Year effects	State effects[&]	Test Score in levels[a]
Cumulative Formal Child Care	0.00302	0.00241	0.00290	0.00299	0.00266	-0.00157	0.36834
	(0.0066)	(0.0064)	(0.0085)	(0.0068)	(0.0068)	(0.0064)	(0.5131)
Cumulative Informal Child Care	-0.00643	-0.00721	-0.00867	-0.00644	-0.00610	-0.01013	-0.48528
	(0.0029) **	(0.0029) **	(0.0035) **	(0.0029) **	(0.0030) **	(0.0061) *	(0.2312) **
Log(Cumulative Income)	0.00719	0.00717	0.05873	0.00732	-0.00054	0.01755	0.25614
	(0.0233)	(0.0249)	(0.0261) **	(0.0237)	(0.0259)	(0.0288)	(1.8241)
Mother's education	0.01203	0.01381	0.01495	0.01200	0.01326	0.01359	0.92806
	(0.0032) **	(0.0034) **	(0.0038) **	(0.0033) **	(0.0032) **	(0.0036) **	(0.2549) **
Mother's AFQT score	0.00061	0.00060		0.00061	0.00062	0.00066	0.07486
	(0.0003) **	(0.0003) **		(0.0003) **	(0.0003) **	(0.0003) **	(0.0259) **
Child's age	0.04067	0.04173	0.02843	0.04068	0.04326	0.04583	3.58776
	(0.0123) **	(0.0126) **	(0.0137) **	(0.0123) **	(0.0127) **	(0.0139) **	(0.9379) **
Mother's age		-0.01208	-0.00696				
		(0.0156)	(0.0166)				
Mother's age squared		0.00022	0.00007				
		(0.0003)	(0.0003)				
I[age of mother$_i$<20]	0.02016	0.00535	0.00607	0.01986	0.00924	0.01999	1.45023
	(0.0114) *	(0.0153)	(0.0162)	(0.0142)	(0.0117)	(0.0113) *	(0.9125)
I[age of mother$_i$>=33]	0.01236	0.00263	0.01098	0.01227	0.00801	0.01536	0.88193
	(0.0254)	(0.0322)	(0.0353)	(0.0256)	(0.0258)	(0.0259)	(2.1840)
Mother's age at first birth		-0.00163					
		(0.0018)					
Number of children	-0.02471	-0.02517	-0.02115		-0.02026	-0.02937	-2.00045
	(0.0061) **	(0.0058) **	(0.0073) **		(0.0064) **	(0.0079) **	(0.4889) **
Number of children 0-5				-0.02458			
				(0.0066) **			
Number of children 6-17				-0.02497			
				(0.0103) **			
Year (at time of test)					-0.01052		
					(0.0038) **		
Year squared					0.00056		
					(0.0003) **		
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	3787	3787	3787	3787	3787	3787	3787
R-squared	0.3776	0.3743	0.3318	0.3776	0.3828	0.3649	0.3916
k [#]	1.039	1.037	1.047	1.038	1.041	1.040	1.036
Weak/Many Instruments Test	4.50	4.61	4.55	4.42	4.50	4.15	4.50

Instruments are all 78 policy variables, local demand conditions and interactions described in footnote in Table 5 in the main text.

[a] The mean score is 91.9, so the point estimate implies an informal child care effect of -0.53 per quarter, or -2.1% per year.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

[&] Joint-significance Test for State F.E.: 28.25 (0.104)

Robust standard errors (Huber-White) by child clusters.

** Significant at 5%; * Significant at 10%

Web Appendix I: Table I2
Robustness with respect to mother's age at childbirth

Dependent Variable -> Log(Test Score)

	(Mother's) Age Restricted Samples		
	24+	24-34	24-30
Cumulative Formal Child Care	0.00064	0.00071	0.00307
	(0.0067)	(0.0062)	(0.0086)
Cumulative Informal Child Care	-0.00720	-0.00600	-0.00466
	(0.0050)	(0.0046)	(0.0059)
Mother's education	0.01102	0.01060	0.01023
	(0.0048) **	(0.0048) **	(0.0063)
Mother's AFQT	0.00052	0.00045	0.00018
	(0.0005)	(0.0005)	(0.0006)
Estimation Method	LIML	LIML	LIML
Number of Observations	1,680	1,643	1,345
R-squared	0.3714	0.3764	0.3859
k [#]	1.080	1.074	1.096
Weak/Many Instruments Test	4.03	4.16	3.62

Instruments are all 78 policy variables, local demand conditions and interactions described in footnote in Table 5 in the main text.

Robust standard errors (Huber-White) by child clusters.

[#] k is the parameter of the k-class estimator.

** Significant at 5%, * Significant at 10%

Web Appendix I: Table I3

A. Robustness with respect to the Instrument List

Dependent Variable -> Log(Test Score)

	(1) Original set of IVs [&]	(2) Excluding CCDF [a]	(3) Only TL, WR and ED	(4) Excludes TL, WR and ED	(5) Only BEN and local demand	(6) Original set without interactions	(7) Only State-specific instruments [b]	(8) Only State-specific IVs [b] without interactions
Cumulative Formal Child Care	0.00302	0.00268	0.00067	-0.00208	-0.12307	0.00335	0.00367	0.00572
	(0.0066)	(0.0066)	(0.0067)	(0.0761)	(0.5455)	(0.0063)	(0.0060)	(0.0126) *
Cumulative Informal Child Care	-0.00643	-0.00713	-0.00946	-0.00439	0.01311	-0.00617	-0.00765	-0.00719
	(0.0029) **	(0.0029) **	(0.0041) **	(0.0097)	(0.0792)	(0.0028) **	(0.0026) **	(0.0033) **
Log(Cumulative Income)	0.00719	0.00697	0.02390	0.01842	0.11612	-0.00439	-0.01152	-0.01400
	(0.0233)	(0.0239)	(0.0355)	(0.0767)	(0.5549)	(0.0230)	(0.0251)	(0.0425)
Mother's education	0.01203	0.01232	0.01252	0.01035	0.01783	0.01180	0.01270	0.01096
	(0.0032) **	(0.0032) **	(0.0034) **	(0.0053) *	(0.0303)	(0.0031) **	(0.0034) **	(0.0037) **
Mother's AFQT	0.00061	0.00062	0.00056	0.00056	0.00100	0.00068	0.00075	0.00074
	(0.0003) **	(0.0003) **	(0.0003) **	(0.0004)	(0.0016)	(0.0003) **	(0.0003) **	(0.0004) **
Estimation Method	LIML	LIML	LIML	LIML	LIML	LIML	LIML	LIML
Number of Observations	3,787	3,787	3,787	3,787	3,787	3,787	3,787	3,787
R-squared	0.3776	0.3735	0.3576	0.3861	-	0.3766	0.3634	0.3568
k [#]	1.039	1.038	1.034	1.005	1.002	1.007	1.024	1.005
Weak/Many Instruments Test	4.50	4.59	5.13	0.81	0.44	5.91	3.90	2.67
P-value, Overidentification test	0.648	0.615	0.392	0.968	-	0.894	0.786	0.803
Number of instruments	78	75	58	27	18	26	63	25

[&] All 78 policy variables, local demand conditions and interactions described in footnote in Table 5 in main text. Unless otherwise noted in column heading, all specifications still include these in

[a] See descriptions of instruments in Table 1. CCDF: ChildcareDevelopment Fund expenditures; TL: time limits; WR: work requirements; ED: earnings disregards; BEN: benefit amounts.

[b] Excludes all individual-specific welfare rules, such as, whether a woman could have a hit a time limit or a work requirement (i.e., all variables with an *i* subscript in Table 1).

Robust standard errors (Huber-White) by child clusters.

[#] k is the parameter of the k-class estimator, which equals 1 for 2SLS and exceeds 1 for LIML.

** Significant at 5%; * Significant at 10%

B. Explanatory Power of Instruments in First Stage
Regressions for Informal Childcare (Instruments in Table I3A)

Dependant variable-> Cumulative Informal Childcare

Instruments listed in footnotes in Table F3A	Partial correlation squared	Shea partial correlation squared	Incremental R^2	F-statistic	P-value
Original set of IVs	0.1337	0.1446	0.0921	16.310	0.000
Excluding CCDF	0.1338	0.1391	0.0885	17.270	0.000
Only TL, WR and ED	0.1049	0.0914	0.0694	21.730	0.000
Excludes TL, WR and ED	0.0753	0.0751	0.0499	7.610	0.000
Only BEN and local demand	0.066	0.0387	0.0437	9.830	0.000
Original set without interactions	0.0997	0.1037	0.0714	9.990	0.000
Only State-specific IVs	0.1129	0.1340	0.0747	7.580	0.000
State-specific IVs without interactions	0.0769	0.0899	0.0562	5.760	0.000

R^2 of first stage regression with only exogenous variables=0.3885

Web Appendix I: Table J1
Who is using formal child care and care provided by non-relatives?

Dependent Variable ->	1 if formal childcare used (0 if informal)		1 if care provided by non-relative (0 if relative)	
Mother's education	0.12126		0.11247	
	(0.0149)	**	(0.0167)	**
Mother's age at birth	-0.01140		0.02024	
	(0.0056)	*	(0.0061)	**
Number of children	-0.08925		-0.04318	
	(0.0191)	**	(0.0213)	**
Urban/rural	0.17590		0.63539	
	(0.0637)	**	(0.0798)	**
No. of observations	12,167		9,471	
Method of Estimation	Logit		Logit	
Pseudo R-squared	0.0116		0.0209	